THE SELECTION PROBLEM FOR CONSTITUTIVE PANPSYCHISM

Philip Woodward

Forthcoming in *Australasian Journal of Philosophy*

Abstract

Constitutive panpsychism is the doctrine that macro-level consciousness—that is, consciousness of the sort possessed by certain composite things such as humans—is built out of irreducibly mental (or proto-mental) features had by some or all of the basic physical constituents of reality. On constitutive panpsychism, changes in macro-level consciousness amount to changes in either the way that micro-conscious entities 'bond' or the way that micro-conscious qualities 'blend' (or both). I pose the 'Selection Problem' for constitutive panpsychism: the problem of explaining how high-level functional states of the brain 'select' micro-conscious qualities for bonding or blending. I argue that there are no empirically plausible solutions to this problem.

Keywords

Panpsychism, Neuroscience, Consciousness

1. Introduction

Constitutive panpsychism is the doctrine that macro-level consciousness—that is, consciousness of the sort possessed by certain composite things such as humans—is built out of irreducibly mental (or proto-mental) features had by some or all of the basic physical constituents of reality.

Constitutive panpsychism is motivated by a pair of considerations.[1] First, contrary to standard forms of physicalism, phenomenal properties are irreducible to the structural-dynamical properties that comprise the explanatory apparatus of the physical sciences. Second, contrary to standard forms of dualism, nature is unified and causally closed, amenable to transparent, bottom-up explanation at all ontological levels. Constitutive panpsychism presents an elegant alternative to both physicalism and dualism. Just as the structural-dynamical features of the basic physical constituents of reality explain the structural-dynamical nature of reality at every ontological level, so the intrinsic features of these entities—specifically, their phenomenal or proto-phenomenal features (henceforth, I'll elide the second option)—explain the phenomenal nature of reality at every ontological level (that has a phenomenal nature).[2]

In this paper, I argue that constitutive panpsychism cannot explain the dynamics of consciousness. More precisely, it is unable to explain how distinct sets of micro-conscious qualities (that is, the phenomenal features of the ultimate physical constituents) are 'selected', moment-to-moment, to form an evolving stream of macro-consciousness. I call this 'the Selection Problem'.

2. Getting Constitutive Panpsychism Up and Running: Bonding and Blending

Constitutive panpsychists notoriously face the 'Combination Problem,' the problem of explaining how micro-phenomenal 'ingredients' could combine to form consciousness as we

[1] See, for example, the 'nonreductionism' and 'nonemergence' premises in Nagel [1979]; the 'argument from intrinsic natures' and the 'genetic argument' discussed by Bruntrup and Jaskolla [2017]; the 'conceivability argument' and the 'causal argument' in Chalmers [2017a].

[2] The view I am calling 'constitutive panpsychism' could also be called 'constitutive *micro*psychism', since it includes an explicit commitment to 'bottom up' explanatory structure. It thus contrasts with constitutive *cosmo*psychism, a version of panpsychism committed instead to *top-down* explanatory structure. I return to the topic of constitutive cosmopsychism in the concluding section below.

know it. Chalmers [2017b] has anatomized the Combination Problem, revealing several sub-problems, including:

- The subject-summing problem: the problem of explaining how a unitary macro-subject (that is, a bearer of macro-conscious states) can arise from a plurality of micro-subjects (that is, the bearers of micro-conscious states);
- The palette problem: the problem of explaining how the multiplicity of types of macro-conscious qualities, and the multiplicity of determinates within those types, can arise from the qualities supposedly instantiated at the micro-level.

I will now sketch the picture that emerges from the literature on the subject-summing and palette problems, since these details will be relevant later on.

The subject-summing problem derives from the explanatory gap that holds between aggregates of micro-subjects, on the one hand, and the macro-subjects they are said to constitute, on the other. The simplest solution to the problem is to espouse a form of mereological universalism for micro-subjects. But this 'solution' to the problem is unattractive because it implies that the macro-subject I call 'me' is no less unified than a 'subject' consisting of seven quarks scattered throughout the universe; with the fusion of me and all the quarks making up the Orion Nebula; and so on. There must be something more to subject-summing than mere aggregation, in other words.

Goff [2017a] points out that our physical theories make mention not only of entities and their monadic properties, but also of relations that hold among the entities. But the physical theories themselves only inform us about the structural-dynamical side of reality, not its intrinsic nature, and this is no less true of relations than of entities and monadic properties. Goff proposes

that the intrinsic nature of one of the physical relations does the work of uniting micro-subjects. He calls this hidden relational property 'phenomenal bonding'.

But *which* of the physical relations has this intrinsic nature? Goff tentatively proposes the *spatial* relation as the relation in question. Roelofs [2019] suggests that the imposition of one of the physical *forces* might play the bonding-role.[3] A third alternative, discussed by both Chalmers [2017b] and Goff [2017b], is that the bonding-relation is a neurologically-specific causal relation, such as being-a-part-of-a-brain-like system, or being-neurally-informationally-related in certain ways. For convenience, I'll extend Goff's label 'phenomenal bonding' to cover all three of these options.

The palette problem derives from the observation that macro-conscious states contain myriad qualities native to multiple, discreet quality-spaces. For example, not only can conscious experiences include any number of determinate *color*-qualities, they can also include any number of determinate *olfactory* qualities, and qualities of the two types are apparently irreducible to one another. Whence this rich array of macro-conscious qualities?

There are two theoretical choice-points confronting any putative solution to the palette problem. The first is the choice between positing many or few basic micro-phenomenal qualities. Chalmers [2017b] calls the two theoretical options 'small-palette' and 'large-palette' solutions, respectively. Constitutive panpsychists are apt to prefer small over large palettes, for three reasons. The first is parsimony: all things considered it would be better to posit a few basic properties than a great many. The second is fungibility: if two structures are *physically* identical to each other, then they can be swapped out for one another in an organism without any shift in function, including mental function. (This principle rules out a view on which some up-quarks

[3] Roelofs calls the relevant relation 'phenomenal unity', whose nature is introspectively manifest, he claims, rather than 'hidden', per Goff.

have a particular shade of red as their intrinsic nature whereas other up-quarks have the sound of pure middle-C as their intrinsic nature, and so on.) A third motivation is *Russellianism*, according to which each fundamental physical property (mass, charge, spin, and perhaps others) has a proprietary 'quiddity' as its categorical base. According to Russellian constitutive panpsychism, there are only as many micro-phenomenal qualities as there are fundamental physical properties.

This is not to say that large-palette solutions are ruled out for the constitutive panpsychist. Lewtas [2017] offers two ways that a large-palette solution can respect fungibility (though not parsimony or Russellianism). On the first, all basic physical entities of a single type instantiate myriad phenomenal qualities; when these entities partly compose certain complex physical structures, one of these qualities gets 'chosen' as the property experienced by that structure as a whole. On the second, all basic physical entities of a single type instantiate a determinable property corresponding to one of the phenomenal quality-spaces (color, smell, sound, etc.); when these entities partly compose certain complex physical structures, this determinable property gets 'collapsed' into a determinate quality experienced by that structure. On both of these proposals, it is the structure of the composite entity of which a basic physical entity is a part that determines which micro-quality that entity contributes to the macro-palette.

The second choice-point has to do with the nature of the relation that holds between micro- and macro-conscious qualities. Given the richness of the macro-palette, only proposals that posit exceedingly large micro-palettes will be able to think of this relation as *identity*. For all others, the relation in question will involve some form of 'blending'. The best-developed account of blending comes from Roelofs [2014]. On Roelofs' proposal, blending occurs when multiple micro-conscious qualities are experienced together. He illustrates the idea by appealing to cases

in which an aesthetic novice treats certain phenomenal qualities as simple that a connoisseur can recognize as complex—the sound of a musical triad or the taste of wine, for example. Roelofs calls this inability of ours to distinguish co-experienced micro-conscious qualities 'confusion'.[4] Lewtas [2017] describes a related proposal, on which blending is akin to a pointillist illusion.

There are thus three ways that a constitutive panpsychist can go about solving the palette problem: an extreme large-palette solution that posits one micro-quality for every determinate macro-quality; a moderate large-palette solution, where a composite entity determines which micro-quality each of its basic physical constituents contributes to the macro-palette, via Lewtas-style 'choosing' or 'collapse'; a small palette solution in tandem with Roelofs-style 'confusion'. Because the first option violates fungibility, I'll set it aside, and for convenience I'll extend Roelofs' label 'phenomenal blending' to cover choosing, collapse and confusion.

3. The Selection Problem

Solving the Combination Problem in its various forms, including the two just discussed, does not fully clear the way for constitutive panpsychism: it still faces the Selection Problem. The Selection Problem emerges from the fact that consciousness changes. This fact is among the most easily demonstrated of facts: simply blink, or move your head, and you have witnessed a change in your (visual) consciousness. Such changes in consciousness are a matter, for the constitutive panpsychist, of changes with respect to which qualities are 'selected' for inclusion in a bonded/blended macro-consciousness. The Selection Problem is the problem of explaining how

[4] Note that Roelofs takes confusion to be the default result of bonding. Thus, blending is to be explained in terms of the absence of a differentiation-mechanism rather than the presence of blending-mechanism. More on this in section 3 below.

such ‘selection’ occurs, consistent with what neuroscience teaches about the architecture and dynamics of the brain.[5]

A brief review of the relevant empirical details is in order. Like everything else, brains are made of fundamental stuff. In particular, they are made of up-quarks, down-quarks, and electrons. Two up-quarks and a down-quark joined together via the strong force constitute a proton; two down-quarks and an up-quark joined together via the strong force constitute a neutron. Six protons plus between six and eight neutrons all joined together via the weak force constitute a carbon nucleus.

Like everything else in the body, brains are made predominantly of carbon-based molecules—in which hydrogen, oxygen, and nitrogen are carbon’s most common bonding-partners—of four types: fatty acids, sugars, amino acids, and nucleotides. These molecules make up the vast majority of the chemical structures and mechanisms found in cells.

The brain is a composite of cells of two types (plus blood vessels, fluid, and other types of packaging): neurons and glia, both of which admit of many sub-types. While glial cells perform many important functions in the brain—not all of which are well understood—most brain research continues to be focused on neurons and their functions, since glial activity facilitates neural activity rather than the other way around, and neural activity seems to be at the heart of what the brain does. (More on this below.) What makes neurons so special are the branches that project outward from their cell body, known as ‘dendrites’ and ‘axons’, which are generally thought of as signal-receiving and signal-sending branches, respectively. All cells are

[5] Einar Duenger Bohn [2018] has recently proposed an alternative to constitutive panpsychism that he calls ‘pluralized panpsychism’. On this view, macro-consciousness is a fundamental property exhibited by a plurality of basic physical entities. He contends that the Combination Problem does not come up for his view. It is worth noting that the Selection Problem *does* still come up: we are still owed an explanation of how the physical dynamics of the plurality bring about changes in that fundamental property.

enclosed by a 'skin' made predominantly of lipids (structures made of fatty acids) and proteins (structures made of amino acids). One function of these cell membranes is to keep negatively-charged ions outside of cells and to keep positively-charged ions inside of cells. The membrane enclosing neuronal dendrites and axons has an added feature: it is punctuated by so-called 'voltage gated ion-channels', that is, apertures that allow for the movement of molecules in and out of the cell. When these gates open, positively-charged sodium ions flow into the cell from outside. The resulting cascade of electrical de-polarization down the length of an axon is called an 'action potential'. When such a cascade reaches the axon's extremity, it occasions the transmission of 'neurotransmitter' compounds across a gap ('synapse') between the tip of the axon and the dendrite of another neuron.

Via synaptic connections, the neurons in the brain form an immensely complicated, interconnected functional web. One neuron's action potential can cue action potentials virtually anywhere else in the brain. Feedback loops make possible non-linear effects including, among other things, synchronous action-potentials within assemblies of many neurons acting as functional units. Coordinated activity within anatomically localizable assemblies correlates with (subjects' reports of) specific types of changes in consciousness. For example, there are assemblies of neurons in visual cortex whose firing correlates with subjects' conscious experience of certain colors, shapes, motions, and so forth. Activations of such assemblies (or sometimes the assemblies themselves) that seem to 'code for' particular conscious qualities are referred to as 'neural correlates of consciousness'.[6]

In light of these details about the brain, we can frame the Selection Problem informally as follows. If constitutive panpsychism is correct, then there has to be some way for

[6] For details and ongoing controversies related to the neural correlates of consciousness, see, e.g., Koch [2004], Tononi and Koch [2008], Hohwy and Bayne [2015].

bonding/blending between and among the relevant micro-conscious physical structures to undergo changes that are explanatorily downstream from activity in specific neural assemblies. But there is no apparent way for this to occur.

More formally:

1. If constitutive panpsychism is true, changes in consciousness (in beings like us) are wholly grounded in changes with respect to phenomenal bonding among micro-subjects or phenomenal blending among the qualities had by micro-subjects (for short: changes with respect to bonding/blending).
2. Changes in consciousness (in beings like us) are explanatorily dependent on changes with respect to high-level, global, dynamical properties of brains.
3. For any distinct phenomena A and B, if phenomenon A is explanatorily dependent on phenomenon B, then whatever wholly grounds phenomenon A is explanatorily dependent on (or identical to) phenomenon B.
4. So, if constitutive panpsychism is true, changes with respect to bonding/blending are explanatorily dependent on changes with respect to high-level, global, dynamical properties of brains.
5. If changes with respect to bonding/blending are explanatorily dependent on changes with respect to high-level, global, dynamical properties of brains, there is a physical, causal mechanism in the brain that inputs changes with respect to high-level, global, dynamical properties and outputs changes with respect to bonding/blending.
6. There is no purely physical, causal mechanism with this profile.
7. So, constitutive panpsychism is false.

I take premise 3 as a fairly obvious principle about explanatory dependence. In case it does not strike the reader as obvious, I offer the following illustration as motivation. A choir sounds a unison note with a particular pitch, timbre, volume, and duration. The sounding of the note is wholly grounded in respiratory-acoustic events occurring within and around the bodies of the choristers. No viable explanation of the note's character is explanatorily independent of these respiratory-acoustic events. And this is true whether the explanans is wholly metaphysically distinct from the grounds—for example, the shape of the hall, the motions of the conductor—or not wholly distinct—for example, the velocity of exhalation, the number of choristers. All viable explanations will be 'routed through' the relevant respiratory-acoustic events. Premise 3 generalizes from this observation. It says that that nothing can explain a phenomenon A while bearing no explanatory relationship to A's grounds.

So much for Premise 3. Premise 4 is a sub-conclusion that rests on premises 1-3. So, the substantive premises are 1, 2, 5, and 6. I now discuss each of these premises in detail.

Premise 1. *If constitutive panpsychism is true, changes in consciousness (in beings like us) are wholly grounded in changes with respect to phenomenal bonding among micro-subjects or phenomenal blending among the qualities of micro-subjects (for short: changes with respect to bonding/blending).* This premise assumes that constitutive panpsychists must appeal to theories of phenomenal bonding and phenomenal blending in order to solve the subject-summing and palette problems, respectively. Perhaps there are ways to solve, or dissolve, those two problems while rejecting this assumption. I grant that my argument has no force against these alternatives; but in light of the recent literature discussed in section 2 above, the assumption seems well founded.

There is more than one way for a constitutive panspsychist to understand the relationship between bonding and blending. On one view of the relationship, the architecture of the brain guarantees more-or-less stable bonding among some set of micro-subjects that make it up, but within this stable composite, changes occur regarding the way micro-qualities blend. Call this the 'bonding-first' paradigm. We can contrast this with the 'blending-first' paradigm, on which the architecture of the brain guarantees more-or-less stable blending within certain chemical structures, but among these structures, changes occur regarding which structures bond with one another.

There are more complicated options. Perhaps blending occurs at two levels: stably, as a prerequisite to bonding, and then differentially, within the bond. Perhaps neither bonding nor blending is stable, but both are independent ways that macro-consciousness can change. For simplicity, I will focus my discussion on the two simplest options. My criticisms will extend to more elaborate theories of the hierarchy of composite consciousness.

Note that I am restricting my discussion to *qualitative* changes in consciousness. There may be changes that outstrip the qualitative: for example, changes pertaining to semantic structure (that is, which qualities are co-attributed); attentional shifts; changes in the 'level' of consciousness (that is, whether one is wide awake, sleepy, dreaming, etc.). Though interesting in their own right (and in need of explanation by constitutive panpsychists), such non-qualitative changes are only relevant to the Selection Problem if they entail qualitative changes (as, for example, Roelofs maintains—see the discussion of premise 5 below).

Premise 2. *Changes in consciousness (in beings like us) are explanatory dependent on changes with respect to high-level, global, dynamical properties of brains.* When I say that certain changes in conscious states are 'explanatorily dependent' on changes in brains, I mean

something neutral among various dependence-relations, including identity (the limiting case of explanatory dependence), constitution, causation, nomic necessitation, or something even weaker,[7] though I argue below that constitutive panpsychists will need to think of such dependence in causal terms.

I emphasize that the dependence is on *dynamical* properties of brains. Perhaps brains also undergo changes with respect to intrinsic qualities that are not identical to dynamical properties, but I am interested in those properties that neuroscience treats as explanatorily relevant.

Why think that changes in conscious states are explanatorily dependent on changes with respect to *high-level* and *global* properties of the brain, rather than with respect to local and/or low-level (physical or chemical) properties of the brain? That the relevant properties are *high-level* is evident in two ways. First, this is how neuroscientists talk. They talk of the 'firing' or 'spiking' of individual neurons, or of 'signaling' among neural assemblies. None of these ways of talking mentions the physical and chemical mechanisms that make action potentials possible. Rather, neuroscientists understand that neurons have an essential *function*—that is, a critical role in the economy of an organism—just as hearts and other organs do. This function can be specified, and to a certain extent studied, while abstracting away from the details of the mechanisms that implement it. Second, and relatedly: the series of events that implements this function of sending electrical signals is strikingly heterogeneous: the opening of a series of gates in the membrane, the motion of ions, the release of neurotransmitters, and the chemical bonding of these proteins to neuroreceptors. These events are unified only in so far as they jointly realize a high-level functional property.

[7] As an example of something weaker, in Woodward [2019] I suggest that the relationship is something like an *interpretation*-relation.

Why think that neural correlates of changes in consciousness are *global* properties of the brain? Suppose Koch [2004] is right that activity in the section of visual cortex known as 'V4' correlates with visual phenomenology. While it is tempting to think of 'activity in V4' as a local (albeit high-level) property of the brain, we should resist this temptation, because what it is to *be* V4 is not a local matter. This is evident, first, from the fact that neurons in V4 have no intrinsic features that makes them especially suited to represent color-stimuli (say). What makes it the case that their activity represents color-stimuli is their particular place in the informational network of the brain. Second, the possibility of variation across brains, and in the same brain over time (so-called "neural plasticity") means that gross anatomy is only a contingent guide to functional specification in cortex. An extreme example is a girl who was born with only one brain hemisphere, but who has developed more or less normally [Muckli, Naumer, and Singer 2009].

In short, the neural events that drive the dynamics of consciousness high-level functions in more or less the whole brain (or at least a major chunk of the cortex and thalamus).

Premise 5. *If changes with respect to bonding/blending are explanatorily dependent on changes with respect to high-level, global, dynamical properties of brains, there is a physical, causal mechanism in the brain that inputs changes with respect to high-level, global properties and outputs changes with respect to bonding/blending.* Given the innocuous principle about explanatory dependence stated in premise 3, it follows that if constitutive panpsychism is true, changes with respect to bonding/blending are explanatorily dependent on changes with respect to high-level, global, dynamical properties of brains. Constitutive panpsychists need a story to tell about this explanatory relationship. Premise 5 specifies that the story needs to appeal to *causation* to account for this explanatory dependence, rather than to a tighter dependence-

relation such as constitution, or a looser dependence-relation such as nomic necessitation. Why can't constitutive panpsychists say that changes with respect to neural firing *constitute* changes with respect to bonding/blending? Consider, by way of analogy, a rotating bicycle wheel. The spokes of the wheel change their orientation vis-à-vis the surface of the road, from perpendicular to oblique to parallel and back again, and these changes are explanatorily dependent on the spinning of the wheel. There is nothing mysterious about such dependence, since changes in the orientation of the wheel *constitute* changes in the orientation of its parts. Constitutive panpsychists could say something similar, by identifying some part of the neural-firing process with changes in bonding/blending.

Exactly how such a story could go depends on whether we're working within a bonding-first or a blending-first paradigm. Let's start with a blending-first paradigm, on which changes in macro-consciousness amount to changes in phenomenal bonding. Extant theories identify phenomenal bonding with (A) spatial relations, (B) relations of physical force imposition, or (C) high-level causal relations such as *being part of a brain* or *being neurally-informationally related*. Are there entities in the brain such that neural firing makes a constitutive difference to which other entities they bear these relations to?

The answer is yes. (A) Action-potentials involve changes with respect to the spatial relations that hold between, for example, sodium ions and the axon membrane they move across. (B) Synaptic transmission involves changes with respect to the chemical bonds between neurotransmitters and neuroreceptors. (C) There is no clear sense in which the firing of neural assemblies alters which entities are part of the brain, but such firings do alter the informational relationships among neural assemblies that code for different conscious qualities.

For any of the relational changes just mentioned to explain changes in phenomenal bonding, there would have to be a one-to-one matching between (a) the firing of a neural assembly that 'codes for' some particular conscious quality Q, and (b) changes in relations (A)-(C) among structures of the proprietary type that exemplify the blended quality Q. But there is no such one-to-one matching. The structures involved in changes (A)-(C) are *neuronally generic*—they are of a range of structure-types whose distribution across the brain is independent of local coding-function. Thus, the firing of a particular neural assembly brings about relational changes among structures of many types, and the firing of any number of neural assemblies brings about relational changes among structures of a single type. Hence, if phenomenal bonding is understood in ways (A)-(C), the firing of neural assemblies cannot provide us with a constitutive explanation of changes in phenomenal bonding. Moreover, any novel suggestion about the nature of phenomenal bonding—beyond (A)-(C)—will be subject to the same worry about a lack of neuronal specificity among the structures constitutively affected by neural firing.

Within a bonding-first paradigm, changes in macro-consciousness amount to changes in blending among stably-bonded entities. Extent theories identify phenomenal blending with (D) the 'collapse' or 'choosing' of micro-conscious qualities when basic physical entities combine to form chemical composites, and (E) 'confusion', that is, indistinguishability among micro-phenomenal qualities that are experienced together. Does the firing of neural assemblies create changes with respect to which basic physical entities are related in these ways?

Let's consider each suggestion in turn. (D) Neural firing does involve the formation of new chemicals. In particular, the processes that precede the action potential—the neural 'computations' that determine whether, on the basis of excitatory and inhibitory signals received by the neurons' many dendrites, the neuron feeds the signal forward—are quite chemically

complicated and require the generation of new molecules.[8] The trouble, as before, is that these chemical changes are neuronally generic: there are no special chemicals generated in the ramp-up to the action potential only in neurons within assemblies that code for particular macro-conscious qualities. Thus, the chemical changes required for neural firing cannot be those responsible for changes in blending.

(E) In order to tell whether neural firing could alter which qualities are 'confused' with each other, we first need a neurological account of the phenomenon of confusion. I will focus on Roelofs' [2019] account, although the lesson to be drawn from his account generalizes. According to Roelofs, confusion is the default result of phenomenal bonding. Blending, that is, comes 'for free'; qualitative differentiation calls out for explanation. Roelofs suggests that what explains differentiation is *semantic structure*—the sort of structure exhibited by redness and squareness when these qualities are co-attributed to the same gestalt object (a red square). Thus, for Roelofs, changes in consciousness are determined by changes in which (otherwise confused) conscious qualities segregate into distinguishable phenomenal units via semantic binding.

There are two leading theories of the neural correlates of semantic binding.[9] According to one, binding occurs when neural assemblies that code for different conscious qualities fire in synchrony. According to another, there are specific neurons or neural assemblies that code for specific conjunctions of qualities. It would appear, then, that Roelofs' proposal allows us to draw a line of constitutive explanation from the firing of certain neurons or neural assemblies (multiple in synchrony, or single conjunctive-coders) to changes in blending. But Roelofs' proposal is subject to a dilemma. Binding is either an occurrent phenomenon (first horn) or a dispositional phenomenon (second horn). Suppose it is occurrent: it is part of the 'fabric' of

[8] Thanks to Jim Nelson for clarification on this point.
[9] See, e.g., Schneegans and Bays [2017].

consciousness, a relation that holds among qualities in macro-consciousness. If so, it is not clear how neural firing can explain *which qualities* in macro-consciousness are subject to binding at a time. There is, for example, nothing about the intrinsic structure of three neural assemblies, X, Y, and Z, that explains why, when X and Y fire in synchrony with each other but not with Z, a subject experiences semantic binding between phenomenal blue and phenomenal yellow, but experiences phenomenal red as a distinguishable quality. Again: neural assemblies are all made of the same, neuronally-generic building-blocks.

Perhaps Roelofs understands binding as a dispositional phenomenon, then. In this case, the binding of two qualities in macro-consciousness is a matter of the way they dispose the subject to treat them (that is, as blended or distinct). It is easier to understand how synchronous or conjunctive-coder neural firing could explain particular changes to binding, dispositionally understood; neurons and neural assemblies, though not intrinsically differentiated, *are* functionally differentiated. But the trouble with the second horn is that it implies that the dynamics of consciousness is entirely a dispositional matter. Occurrent consciousness comprises a static bond of qualities; all that changes are the discriminatory dispositions of the subject. But that's absurd. Occurrent consciousness changes; conscious qualities come and go.

This dilemma for Roelofs' account generalizes to *any* constitutive account of the relationship between neural firing and blending. Neurons and neural assemblies are composed of neuronally-generic stuff. The changes they undergo when firing will thus be unable to account for which particular qualities enter into occurrent blending-relations. Neural assemblies are functionally differentiated, however, so their firings may be able to explain changes in a subject's dispositions with respect to her experiences. But changes in consciousness are more than changes in the subject's dispositions.

Constitutive panpsychists are unable to explain the dependence of changes in bonding/blending on high-level properties of brains by appeal to a tighter-than-causal (that is, a constitutive) relationship. What about a looser-than-causal relationship—that is, a merely nomic relationship? Might it just be a law of nature that brains, *qua* electrical signal-sending networks, occasion bonding/blending among their most basic physical constituents in certain circumstances? Maybe it is, but constitutive panpsychists are barred from saying this. As Lewtas puts the point, such a proposal would 'envision basic mental-chemical laws whose antecedents directly refer to high-level structures S1, S2, S3, etc. These structures therefore have *basic* but high-level capacities (high-level capacities that don't result from bottom-level entities)' [Lewtas 2017: 771]. But constitutive panpsychists eschew all emergence—understood as the appearance of fundamental entities, properties, or powers at non-basic ontological levels—in favor of bottom-up explanation. Therefore, the only way for constitutive panpsychists to adequately explain the dynamics of consciousness is to point to a causal mechanism that can mediate between high-level, global, dynamical states and states of aggregates of basic physical entities, and this mechanism needs to be explicable in a fully bottom-up way.

Premise 6. *There is no purely physical mechanism that can ground the explanatory dependence of changes with respect to bonding/blending on changes with respect to high-level, global, dynamical properties of brains.* Broadly speaking, we know of two mechanisms that transfer information within the brain. The first is the familiar mechanism of neural signaling via action potentials. The second is a less familiar mechanism of electromagnetic propagation via field potentials. I'll discuss the three most plausible models that invoke these mechanisms to solve the Selection Problem, plus a fourth, hybrid model.

1. The region-model. Suppose that each region of the brain that codes for a conscious quality *is also the proprietary bearer of that conscious quality*, and suppose that the blending-first paradigm is correct. Neural signals from other parts of the brain cue or disrupt each regions' bonding with the other regions. Call this the 'region-model' of selection.

The trouble with this proposal is that the relata in the bonding-relation need to be bearers of distinct (blended) qualities, so they need to be intrinsically differentiated from one another in some way. But they are not. Though neural assemblies differ from each other in lots of ways, there is no proprietary structure unique to each quality-encoding assembly. As we observed above: what makes it the case that some neural assembly's activation represents some visual stimuli (say) is its place in the informational network of the brain, not its intrinsic structure.

2. The hub-model. While neural assemblies aren't intrinsically differentiated, smaller molecular structures in the brain *are* intrinsically differentiated. Suppose that there is some specific site in the brain where bonding/blending takes place, and suppose (again) that the blending-first paradigm is correct. We can imagine that this 'consciousness hub' consists of a set of distinct structures (differently-shaped proteins, let's say) each of which has a proprietary blended quality, and which can be cued to bond, or to refrain from bonding, with the other proteins at the hub. For simplicity, let's assume that macro-conscious states only require one 'sample' of each blended quality, such that only one instance of each type of structure is needed as part of the bond. There is an electrical signal pathway in the brain that begins at the retina and proceeds through the portion of visual cortex that correlates with the experience of a given color—phenomenal green, for example. Suppose now that the last neuron in this pathway sends an electrical or chemical signal to the consciousness hub: specifically, to the protein at that site whose blended quality is phenomenal green. Upon receiving the signal, our phenomenal-green

structure bonds with other structures that have received similar signals, from their respective processing-channels, at the same time. Call this the 'hub model' of selection.

How empirically plausible is the existence of a 'consciousness hub'? Some researchers have suggested that the prefrontal cortex amounts to a 'convergence zone', where all the processing streams implicated in consciousness (or in some dimension of consciousness) come together. Sevush [2006] claims that it is consistent with our evidence that individual neurons in PFC serve as convergence zones. In particular, Sevush proposes that consciousness is composed of events occurring at the tens of thousands of synapses belonging to dendrites of a single neuron (or rather, many, in parallel) in PFC. But even supposing Sevush is right about this, the synaptic events he appeals to are neuronally generic: they consist of the binding of various neurotransmitters to neuro-receptors. And there is simply no evidence that the different types of neurotransmitter molecules deployed in PFC synapses map onto the different types of macro-conscious qualities.

3. The field-model. The chemical goings-on in neurons (and in glial cells, too) involve a lot of electrical activity. Some of this electrical activity is coordinated enough to generate electromagnetic fields. Most of these 'field potentials' are so weak that they get lost in the electromagnetic 'noise' of the brain within a few millimeters from their source, but some are strong enough to be detected outside the skull (this is what makes EEG possible). Could field potentials do the work of cueing the relevant micro-structures to bond/blend when certain neural assemblies are active? On a bonding-first paradigm, the idea would be that the electromagnetic field generated by the neural assembly that codes for some conscious quality—phenomenal green, say—triggers, within the bonded composite of basic physical entities, a blending process that generates phenomenal green. On a blending-first paradigm, this would mean that the

electromagnetic field generated by the phenomenal-green neural assembly triggers structures of the sort whose proprietary blended quality is phenomenal green to phenomenally bond with the other qualities that are receiving electromagnetic signals from *their* correlative neural areas. Call either of these proposals the 'field model' of selection.

While the field model is biophysically possible, we have pretty strong neuroscientific evidence against it. For one thing, whether a field is generated has more to do with the orientation of a neural assembly and the tissues that surround it than with the structure and function of the assembly itself. Some neural assemblies generate field potentials with comparatively strong amplitudes, while others produce none, and the differences are not systematic. Mainstream neuroscience treats the whole affair as a messy epiphenomenon, rather than a fine-tuned intra-cranial communication system [Herreras 2016].

For another thing, if there were a unique 'signature' (say, a particular frequency) to the field generated by particular neural assemblies, then there would be a way to 'decode' that signature. But this is impossible: researchers are not able to infer the source of a field potential from its characteristics [Wadman and Da Silva 2017]. But the field model only works if electromagnetic signals are differentiated according to the neural assembly that generated them. In short, neural assemblies just aren't set up to broadcast information, by way of field potentials, about the particular qualities they code for.[10]

Might a hybrid model do better? Combining the region-model and the field-model won't get us anywhere, since both models suffer from the same limitation: both of them require that neural assemblies be intrinsically differentiated according to the conscious quality that each codes for. Combining the hub-model with the field-model is more promising. Suppose that

[10] Thanks to Kim Iceman and Oscar Herreras for clarification of these points.

Sevush is right that an entire macro-conscious state could be encoded in the chemical events occurring in the dendritic tree of a single neuron in PFC. If it turned out that (a) these chemical events were sufficiently differentiated to *intrinsically* code for a set of conscious qualities *and* (b) each generated a unique electromagnetic field, *and* (c) there were a microscopic consciousness hub nearby that could pick up these signals, then we could explain selection via a field/hub model.

But none of these three conditions enjoys any empirical support whatsoever. That is, unlike the three models just discussed (hub, region, and field), this hybrid model is not built out of neuroscientifically *given* phenomena, but rather posits new phenomena. True, current neuroscience does not rule out these phenomena. But nor does it rule out more wildly speculative proposals. For example, the constitutive panpsychist could posit a special kind of brain-field that broadcasts information about the high-level behavior of the whole brain to every part of the brain, and to which bonded/blended entities are causally responsive. Perhaps such a field is constituted by a new physical particle, a 'menton' (as Berit Brogaard [2017] has proposed for somewhat different purposes[11]). The point is this: once we move on from models that appeal to neuroscientifically given phenomena such as neural signaling or field potentials, our theories are going to have to be more and more speculative, constituting a greater and greater departure from the trajectory of neuroscience. To the extent that neuroscience is proceeding along the right track, we should be hesitant to take such speculations seriously.

It appears, then, that no purely physical mechanism can explain how changes with respect to bonding/blending can depend on changes with respect to high-level, global, dynamical properties of the brain. It follows that constitutive panpsychism is false.

[11] Peter Unger [2006] credits Yuvul Avner with coining the term.

4. Conclusion: A Strained Explanatory Project

As we have seen, initially promising solutions to the Selection Problem are closed off by the empirical facts. I now want to argue that we should have expected as much. Even a very rough grasp of how the brain relates to consciousness gives us reason to doubt that the Selection Problem is solvable.

Imagine that LEGO® were to produce a line of translucent bricks that light up, in one of the three primary colors, when pressed. Out of these bricks we build a 'marble run'. When we drop a marble into it, we notice that wherever the marble is rolling, that region of the structure lights up—sometimes red, yellow, or blue, but sometimes green, orange, or purple, too. We can easily construct a bottom-up explanation of this phenomenon: the light-colors of individual bricks explain the blended colors of the macro-level parts of structure, and the moving weight of the marble explains why different regions of the marble-run light up at different times.

If the relationship between brain-activity and the dynamics of consciousness were analogous to the relationship between marble-motion and colored-light emission in our marble run, then constitutive panpsychism would look promising. But it is not analogous. What we find in the brain is more like the following. When the marble is in region A of our marble run, the whole LEGO structure emits red light; when the marble is in region B, the whole LEGO structure emits green light; and so on. It is now not so clear how to explain these changes in a bottom-up way. Persisting with constitutive explanation would require saying the following about our LEGO structure: (A) It exhibits system-level qualitative states (its overall color at a time); (B) Changes in these states depend on changes in system-level functional states (which chunk of the structure the marble rolls through at a time); (C) Changes in system-level functional

states are not constituted by changes in micro-level qualitative states (the color of the bricks doesn't determine what the marble is doing); nevertheless, (D) System-level qualitative states (overall color) are constituted by micro-level qualitative states (colors of the individual bricks). That's weird: given (A) – (D), we don't expect there to be a constitutive connection between micro-level qualities and system-level qualities. Insisting on the analogue of (D) would be like insisting that the twisting shapes of a murmuration of birds are constituted by twisting shapes of the animals themselves.

But constitutive panpsychists have to hold out for an analogue of (A) – (D). They are committed to saying that (A) Brains exhibit system-level qualitative states (macro-conscious states); (B) Changes in these states depend on changes in system-level functional states (neural signaling across the brain at a time); (C) Changes in system-level functional states are not constituted by changes in micro-level qualitative states (changes in micro-conscious bonding/blending aren't constituents of the neural signaling process); nevertheless (D) System-level qualitative states are constituted by micro-level qualitative states (macro-conscious states are constituted by bonded/blended micro-conscious states). As in the case of the LEGO marble-run, (A) – (C) suggest that (D) is wrong-headed.

I have been focused entirely on *constitutive* panpsychism, and in particular constitutive *micro*psychism. I want to end by saying a bit about the implications for neighboring views.

First, what about constitutive *cosmo*psychism? According to this view, the universe as a whole is treated as the fundamental tier of reality, and features in every other tier—including all phenomenal features—can be constitutively explained in terms of features of the whole. A choice-point for this view pertains to how much bottom-up explanation it can accommodate. To use Roelofs' helpful terminology: *conservative* cosmopsychism treats the universe as

explanatorily prior to particles, but treats particles as prior to atoms, atoms as prior to molecules, and so forth. *Holistic* cosmopsychists, on the other hand, invert bottom-up explanation at every level. It is thus committed to puzzling claims such as that organisms are prior to cells, cells are prior to molecules, etc.

Because holistic cosmopsychism denies that part-to-whole explanations are *ever* correct, it is hard to take seriously. Conservative cosmopsychism isn't subject to this worry, but for that very reason, the arguments of this paper apply to it. That is, it differs from constitutive *micro*psychism only by positing another tier to reality below that of the basic physical entities (viz., the whole universe). Thus constitutive cosmopsychists face the Selection Problem to the same degree as constitutive micropsychists.

Let's turn now to non-constitutive versions of pan(micro)psychism. Giving up on constitutive explanation opens up new ways to think about (1) bonding, (2) blending, and (3) selecting. By way of illustration: (1) Seager [2017] proposes a non-constitutive account of bonding on which micro-subjects *fuse* with one another to form new, genuinely unified macro-conscious states. (2) Roelofs [2014] discusses a non-constitutive type of blending, also labelled 'fusing', according to which micro-conscious qualities disappear into the state they compose. (3) In connection with Premise 5 above, we considered the possibility that it is just a law of nature (unmediated by any physical mechanism) that brains, qua electrical signal-sending networks, occasion bonding/blending among their ultimate constituents.

As I see it, all of these proposals weaken the central motivation for positing micro-conscious qualities in the first place. For example, a fusion-operation (whether of bonding or blending) on a set of qualities is supposed to generate a state that is ontologically other than, or more than, the mereological sum of those qualities. A natural gloss (perhaps not the only

possible gloss) is that the input-qualities *causally generate* the output-state. But note that the panpsychist's initial reason for positing micro-phenomenal qualities was to *build* macro-phenomenal states out of them. Once we have granted that the explanatory relation between the micro- and the macro- is causal (rather than constitutive), we don't need to posit micro-conscious qualities anymore. We simply need to posit certain causal powers or causal laws had by (collections of) basic physical entities. *Must* the categorical base of a macro-consciousness-generating-power be a phenomenal or proto-phenomenal quality? I don't see why.

Similarly: on a non-constitutive theory of selection, which micro-conscious qualities get selected is a matter of basic law. A panpsychism that embrances an emergentist theory of selection along these lines is committed *both* to new basic laws *and* to micro-conscious qualities. But the dualist emergentist posits basic laws (or basic causal powers), and stops there. *Must* the mechanism for macro-consciousness-generation involve the selection of micro-conscious qualities? I don't see why.

Now, panpsychist emergentisms may enjoy explanatory advantages over dualist emergentisms in quarters I have not here explored. Nevertheless, the move from constitutive to non-constitutive explanation leaves behind one of the two central motivations for panpsychism, and thus amounts to a serious concession.[12] And that means that the Selection Problem is a problem for the panpsychist research program generally, and not just for a parochial version of it.

Acknowledgements

[12] Chalmers [2017b] expresses a similar worry.

Thanks to the following individuals for their helpful comments on early versions of this paper: Kim Iceman, Jim Nelson, David Pitt, Aaron Preston, Luke Roelofs, Stan Zygmunt, and three anonymous referees for this journal.

REFERENCES

Bohn, Einar Dunger 2018. Panpsychism, The Combination Problem, and Plural Collective Properties, *Australasian Journal of Philosophy*. 97/2: 383-94.

Brogaard, Berit 2017. In Search of Mentons: Panpsychism, Physicalism, and the Missing Link, in *Panpsychism: Contemporary Perspectives*, ed. Godehard Brüntrup and Ludwig Jaskolla, New York: Oxford University Press:130-52.

Bruntrup, Godehard and Ludwig Jaskolla 2017. Introduction, in *Panpsychism: Contemporary Perspectives*, ed. Godehard Brüntrup and Ludwig Jaskolla, New York: Oxford University Press: 1-16.

Chalmers, David 2017a. Panpsychism and Protopanpsychism, in *Panpsychism: Contemporary Perspectives*, ed. Godehard Brüntrup and Ludwig Jaskolla, New York: Oxford University Press: 19-47.

Chalmers, David 2017b. The Combination Problem for Panpsychism, in *Panpsychism: Contemporary Perspectives*, ed. Godehard Brüntrup and Ludwig Jaskolla, New York: Oxford University Press: 179-214.

Goff, Philip 2017a. The Phenomenal Bonding Solution to the Combination Problem, in *Panpsychism: Contemporary Perspectives*, ed. Godehard Brüntrup and Ludwig Jaskolla, New York: Oxford University Press: 283-302.

Goff, Philip 2017b. *Consciousness and Fundamental Reality*, New York: Oxford University Press.

Herreras, Oscar 2016. Local Field Potentials: Myths and Misunderstandings, *Frontiers in Neural Circuits* 10/101: 1-16.

Hohwy, Jakob and Timothy Bayne 2015. The Neural Correlates of Consciousness: Causes, Confounds and Constituents, in *The Constitution of Phenomenal Consciousness: Toward a Science and Theory*, ed. Steven Miller, Amsterdam: John Benjamins: 155-76.

Koch, Christoph 2004. *The Quest for Consciousness: A Neurobiological Approach*, Greenwood Village: Roberts and Company.

Lewtas, Pat 2017. Building Minds: Solving the Combination Problem, *Inquiry* 60/7: 742-81.

Nagel, Thomas 1979. *Mortal Questions*. Cambridge: Cambridge University Press.

Lars Muckli, Marcus Naumer, and Wolf Singer 2009. Bilateral Visual Field Maps in a Patient with Only One Hemisphere, *Proceedings of the National Academy of Sciences of the United States of America* 106/31: 13034-39.

Roelofs, Luke 2014. Phenomenal Blending and the Palette Problem, *Thought: A Journal of Philosophy* 3/1: 59-70.

Roelofs, Luke 2019. *Combining Minds: How to Think About Composite Subjectivity*, New York: Oxford University Press.

Schneegans, Sebastian and Paul Bays 2017. Neural Architecture for Feature Binding In Visual Working Memory, *Journal of Neuroscience* 37/14: 3913-25.

Seager, William 2017. Panpsychist Infusion, in *Panpsychism: Contemporary Perspectives*, ed. Godehard Brüntrup and Ludwig Jaskolla, New York: Oxford University Press: 229-48.

Sevush, Steven 2006. Single-Neuron Theory of Consciousness, *Journal of Theoretical Biology*, 238/3: 704-25.

Tononi, Guilio and Christoph Koch 2008. The Neural Correlates of Consciousness: An Update, *Annals of the New York Academy of Sciences* 1124: 239–61.

Unger, Peter 2006. *All the Power in the World*, New York: Oxford University Press.

Wadman, Wytze and Fernando Lopes da Silva 2017. Biophysical Aspects of EEG and MEG Generation, in *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, ed. Donald Schomer and Fernando Lopes da Silva, New York: Oxford University Press: 89-103.

Woodward, Philip 2019. Explaining the Ontological Emergence of Consciousness, in *Consciousness and the Ontology of Properties*, ed. Mihretu Guta, New York: Routledge: 109-25.

Institutional Affiliation

Valparaiso University